Gradient Descent Based Optimization of Transparent Mamdani Systems

Andri Riid and Ennu Rüstern

Tallinn Technical University, Department of Computer Control, Ehitajate tee 5, Tallinn 19086, Estonia. E-mail: andri@dcc.ttu.ee

Abstract. The tradeoff between accuracy and interpretability in fuzzy modeling has shifted into focus in last few years. This paper aims at improving accuracy of linguistic models while maintaining a good interpretability. A new gradient-based method, extended version of Jager approach, is proposed for the optimization of transparent Mamdani systems. The advantage of Mamdani systems if compared to 0^{th} order TS systems in Jager approach is that their interpolation properties allow one to obtain less complex models without loss of accuracy. Several modeling examples confirming the advantages of the chosen algorithm are included.

1 Introduction

Many papers have been written about neuro-fuzzy systems. Most of these address the optimization of fuzzy systems based on gradient descent (GD) method. Perhaps the earliest work in this research line is [1] that considers 0^{th} order TS systems with symmetric triangular antecedent membership functions (MFs). Later contributions include [2] that extends the method on various types of fuzzy systems including 1^{st} order Takagi-Sugeno (TS) systems. The hybrid learning rule, proposed by Jang [3] combines gradient descent with least-squares estimator for consequent parameters and outperforms the previous approaches in terms of approximation error.

What is common to mentioned approaches is that no attention is paid to semantic properties of fuzzy systems, thus the obtained models and controllers are non-transparent to interpretation and it can be claimed that the basic feature that distinguishes fuzzy systems from other nonlinear approximators (e.g. neural networks) is sacrificed to approximation accuracy.

Jager's work where GD is applied to 0^{th} order TS systems in a manner that maintains fuzzy partition on antecedent variables is an interesting exception. His algorithm, published in Ph.D. thesis [4] is, however, not well known with the exception of [5] and [6] (in the latter an extension for 1^{st} order TS systems is given). In this paper we propose a further extension of Jager algorithm for Mamdani systems.

2 System Definition

We consider a MISO Mamdani system with the following rule base

IF X_1 is A_{1r} AND… AND X_i is A_{ir} … AND X_N is A_{Nr} THEN Y is B_{1r}, (1)

where A_{ir} and B_{jr} the linguistic labels of i^{th} input variable x_i and output variable y (i = 1 … N), respectively, associated with the r^{th} rule (r = 1 … R).

Numerical mapping of input-output variables of (1) (in discrete form) is obtained by applying the inference function (2),

$$y = Y_{cog}(F(y)) = \sum_{q=1}^{Q} F(y_q) y_q \Big/ \sum_{q=1}^{Q} F(y_q) \tag{2}$$

where y_q is output value at q^{th} discretization interval and Y_{cog} denotes center-of-gravity defuzzification function.

$F(y)$ in (2) is given by (3), assuming product-product-sum inference.

$$F(y) = \sum_{r=1}^{R} \left(\prod_{i=1}^{N} \mu_{ir}(x_i) \right) \gamma_r = \sum_{r=1}^{R} \tau_r \gamma_r \ , \tag{3}$$

where μ_{ir} and γ_r are normal and convex antecedent and consequent membership functions (MFs), having one-to-one correspondence with respective linguistic labels in (1) and τ_r is the activation degree of the r^{th} rule. (2) and (3) yield

$$y = \sum_{r=1}^{R} \tau_r \sum_{q=1}^{Q} \gamma_r(y_q) y_q \Big/ \sum_{r=1}^{R} \tau_r \sum_{q=1}^{Q} \gamma_r(y_q) \ . \tag{4}$$

According to [7], such system is transparent (i.e. fully interpretable) if antecedent and consequent MFs satisfy (5) and (6), respectively.

$$\forall x_i \in X_i : \sum_{s=1}^{S_i} \mu_i^s(x_i) \le 1 \ , \tag{5}$$

$$\text{Y}_{cog}(\gamma_r(y)) = \int_{y_{\min}}^{y_{\max}} y \gamma_r(y) dy \Big/ \int_{y_{\min}}^{y_{\max}} \gamma_r(y) dy = \text{core}(\gamma_r(y)) \tag{6}$$

If consequent MFs, γ_r, of the system are symmetrical (6) and triangular i.e.

$$\gamma_r(y) = \max\left(0, \min\left(\frac{2b_r - 2y + s_r}{s_r}, \frac{2y - 2b_r + s_r}{s_r} \right) \right), \tag{7}$$

it can be shown [8] that the inference function (4) of a Mamdani system with (7) reduces to (8).

$$y = \sum_{r=1}^{R} \tau_r b_r s_r \Big/ \sum_{r=1}^{R} \tau_r s_r \tag{8}$$

Note that if $\forall s_r = \xi$, where ξ is arbitrary constant, (8) further reduces to the inference function of a 0^{th} order TS system which implies that a Mamdani system with symmetrical triangular consequent MFs is, in fact, a 0^{th} order TS system if the supports s_r are all equal.

To satisfy (5), Jager definition of fuzzy sets [4] is used in the following (a_i^s is the center of s^{th} MF of the i^{th} antecedent variable)

$$\mu_i^s(x_i) = \max\left(0, \min\left(\frac{x_i - a_i^{s-1}}{a_i^s - a_i^{s-1}}, \frac{a_i^{s+1} - x_i}{a_i^{s+1} - a_i^s}\right)\right) \tag{9}$$

3 The Algorithm

Gradient descent method minimizes the objective function E:

$$E = \frac{1}{2}\sum_k^K \varepsilon(k) = \frac{1}{2}\sum_k^K (y(k) - \tilde{y}(k))^2 , \tag{10}$$

where $\tilde{y}(k)$ is the k^{th} reference value and $y(k)$ is obtained by applying the inference function (8) with the k^{th} given input pattern. The update rules based on (8) and (10) have the following form.

$$b_r(l+1) = b_r(l) - \eta_b \frac{\partial E}{\partial b_r(l)} = b_r(l) - \frac{\eta_b}{2}\sum_{k=1}^K \frac{\partial \varepsilon(k)}{\partial b_r(l)} , \tag{11}$$

where η_b is the learning rate of b_r. To compute the derivatives in (10) and similar expressions for other adaptable parameters s_r and a_i^s we apply the chain rule

$$\frac{\partial \varepsilon}{\partial b_r} = \frac{\partial \varepsilon}{\partial y}\frac{\partial y}{\partial b_r} , \tag{12}$$

$$\frac{\partial \varepsilon}{\partial s_r} = \frac{\partial \varepsilon}{\partial y}\frac{\partial y}{\partial s_r} , \tag{13}$$

$$\frac{\partial \varepsilon}{\partial a_i^s} = \frac{\partial \varepsilon}{\partial y}\frac{\partial y}{\partial a_i^s} = \frac{\partial \varepsilon}{\partial y}\left(\frac{\partial y}{\partial \mu_i^{s-1}}\frac{\partial \mu_i^{s-1}}{\partial a_i^s} + \frac{\partial y}{\partial \mu_i^s}\frac{\partial \mu_i^s}{\partial a_i^s} + \frac{\partial y}{\partial \mu_i^{s+1}}\frac{\partial \mu_i^{s+1}}{\partial a_i^s}\right). \tag{14}$$

Once the partial derivatives are obtained, the following update rules can be written:

$$b_r(l+1) = b_r(l) - \eta_b (y(k) - \tilde{y}(k)) \frac{\tau_r(k) s_r(l)}{\sum_{r=1}^R \tau_r(k) s_r(l)} \tag{15}$$

$$s_r(l+1) = s_r(l) - \eta_s (y(k) - \tilde{y}(k))(b_r(l) - y(k)) \frac{\tau_r(k)}{\sum_{r=1}^{R} \tau_r(k) s_r(l)} \tag{16}$$

if $a_i^{s-1} < x_i < a_i^s$ (17)

$$a_i^s(l+1) = a_i^s(l) - \eta_a \frac{(y(k) - \tilde{y}(k))}{a_i^s(l) - a_i^{s-1}(l)} \cdot \frac{1}{\sum_{r=1}^{R} \tau_r(k) s_r(l)} \cdot$$

$$\cdot \left[\frac{\mu_i^s(x_i(k))}{\mu_i^{s-1}(x_i(k))} \sum_{r'=1}^{R(\mu_i^{s-1})} \tau_{r'}(k) s_{r'}(l)(b_{r'}(l) - y(k)) - \sum_{r'=1}^{R(\mu_i^s)} \tau_{r'}(k) s_{r'}(l)(b_{r'}(l) - y(k)) \right]$$

if $a_i^s < x_i < a_i^{s+1}$ (18)

$$a_i^s(l+1) = a_i^s(l) - \eta_a \frac{(y(k) - \tilde{y}(k))}{a_i^{s+1}(l) - a_i^s(l)} \cdot \frac{1}{\sum_{r=1}^{R} \tau_r(k) s_r(l)} \cdot$$

$$\cdot \left[\sum_{r'=1}^{R(\mu_i^s)} \tau_{r'}(k) s_{r'}(l)(b_{r'}(l) - y(k)) - \frac{\mu_i^s(x_i(k))}{\mu_i^{s+1}(x_i(k))} \sum_{r'=1}^{R(\mu_i^{s+1})} \tau_{r'}(k) s_{r'}(l)(b_{r'}(l) - y(k)) \right]$$

where $r' = 1 \ldots R(\mu_i^s)$ refers to rules having A_i^s in their premise.

Note that if $\forall s_r = \xi$, the resulting (15), (17) and (18) constitute the original Jager algorithm for 0^{th} order TS systems.

4 Experimental Results

In the examples below the proposed algorithm is used to model highly nonlinear functions in comparison with ANFIS and the original Jager algorithm.

In order to ensure a good initial configuration of the model, the centers of consequent MFs are determined with a one-step least squares procedure [3], i.e. the model is initialized as a 0^{th} order TS system. The obtained consequent singletons are then converted into symmetrical triangles. To stabilize the learning process, variable learning rate and heuristic step size adjustment strategy [3] are utilized.

First test system is the function

$$y = 0.6\sin(\pi x) + 0.3\sin(3\pi x) + 0.1\sin(5\pi x) \tag{19}$$

Test data set is obtained by discretizing the input variable x = [-1,1] with the step 0.01 that results in 201 training samples. Additionally, two-input single-output fuzzy system from [8], depicted in Fig. 2, right, is identified from 441 training samples with 18, 36 and 66 rules. The results in terms of approximation error obtained with 200 training epochs are depicted in Fig. 1, right and Table 1.

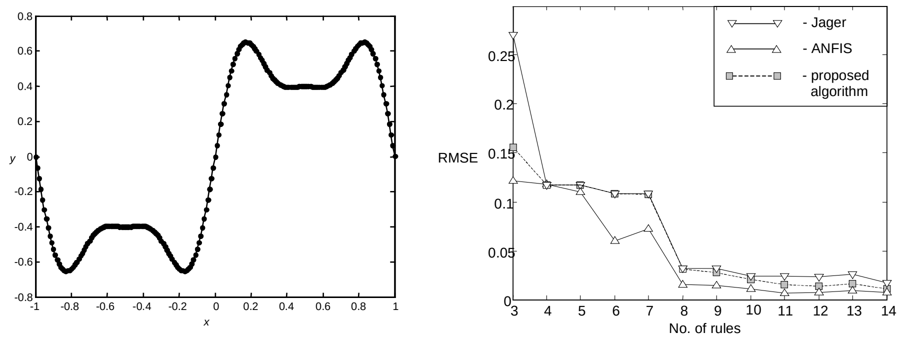

Fig. 1. Training data (left) and modeling root-mean-square-errors (RMSE) (right).

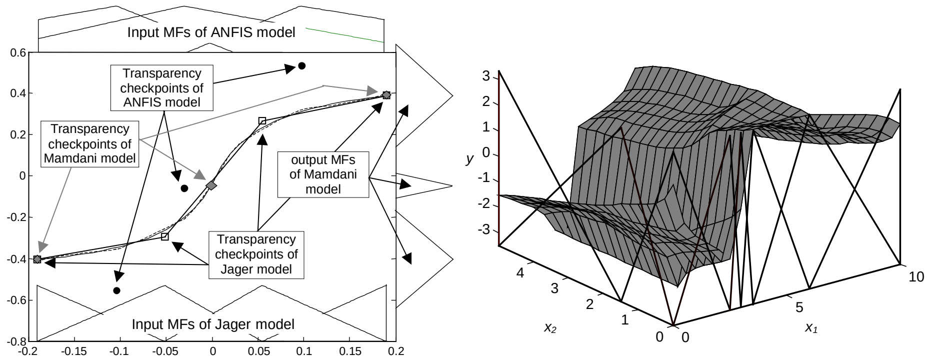

Fig. 2. Interpolation/approximation properties of fuzzy systems (left) and two-input test system (right).

We see that the proposed algorithm consistently outperforms Jager algorithm and while not being real contender to ANFIS makes a good attempt at it. It can be observed, however, that the excellent approximation properties of ANFIS generally come at the expense of transparency. Considering the curve in Fig. 2, left, application of ANFIS results in very low error, Transparency checkpoints (one for each rule, see [8]) that are not situated on the actual system output, however, indicate strong non-transparency. Linguistically such system would be impossible to interpret correctly. Jager's transparent 0^{th} order models, on the other hand, generally require more rules for faithful tracking and even then, linear interpolation does not always provide smoothness. The advantage of a Mamdani system is that while preserving transparency, its interpolation properties allow more efficient approximation.

Table 1. Modeling RMSEs of two input test function.

R	Initial error	ANFIS	Jager	proposed algorithm
6×3	0.5538	0.1594	0.2479	0.1790
9×4	0.2891	0.1178	0.1710	0.1416
11×6	0.1858	0.0780	0.0811	0.0589

5 Conclusions

The neuro-adaptive learning techniques provide a method for the fuzzy modeling procedure to learn information about a data set, in order to compute the membership function parameters that allow the fuzzy system to track the given input/output data. The learning method works similarly to that of neural networks. In present work we have extended the approach onto Mamdani fuzzy systems.

Modeling theory and praxis have proven that approximation error cannot be the only criterion for finding a "best" model. It is theoretically possible to formulate a model that fits any given (finite) learning data set with very high accuracy – according to the rule "the more complicated the model is, the more accurately it will fit the given data." For noisy data, this means that, at a certain point in modeling, the model begins to fit the noise (overfitting), which results in poor performance on new data.

Although overfitting is less potent danger in fuzzy modeling (the parameters of fuzzy systems have physical meaning that serves the same purpose as regularization in neural networks), approximation error still does not provide full information about the "goodness" of the model, because the linguistic properties of fuzzy systems (such as interpretability) and accuracy, are generally contradictory requirements.

The algorithm proposed in this paper attempts to reduce the gap between interpretability and accuracy in fuzzy modeling.

References

1. H. Nomura, I. Hayashi and N. Wakami. A learning method of fuzzy inference by descent method. *Proc. 1st IEEE Int. Conf. on Fuzzy Systems*, pages 485-491, San Diego, CA, USA, 1992.
2. F. Guely and P. Siarry. Gradient descent method for optimizing various fuzzy rule bases. *Proc. 2nd IEEE Int. Conf. Fuzzy Systems*, pages 1241-46, San Francisco, CA, USA, 1993.
3. J.-S.R. Jang. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Trans. Systems, Man and Cybernetics,* 23(3):665-685, 1993.
4. R. Jager. *Fuzzy Logic in Control*, Ph.D. dissertation. Technical University of Delft, Delft, Holland, 1995.
5. J. Abonyi, L. Nagy and F. Szeifert, Indirect Model Based Control Using Fuzzy Model Inversion", *Proc. IEEE Int. Conf. Intelligent Systems*, pages 951-956, Vienna, Austria, 1998.
6. J. Abonyi, H. Andersen, L. Nagy and F. Szeifert. Inverse fuzzy-process-model based direct adaptive control, *Mathematics and Computers in Simulation*, 51(1-2):119-132, 1999.
7. A. Riid and E. Rüstern. Transparent fuzzy systems and modeling with transparency protection. *Proc. IFAC Symp. on Artificial Intelligence in Real Time Control*, pages 229-234. Budapest, Hungary, 2000.
8. A. Riid. *Transparent Fuzzy Systems: Modeling and Control*. Ph. D. dissertation. Tallinn Technical University, Tallinn, Estonia, 2002.